

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-259311

(43)Date of publication of application : 12.09.2003

(51)Int.Cl. H04N 5/93  
G10L 11/02  
G10L 15/00  
G10L 15/02  
G10L 15/04  
G10L 15/06  
G10L 15/10  
H04N 5/91

(21)Application number : 2002-060844 (71)Applicant : NIPPON TELEGR & TELEPH  
CORP <NTT>

(22)Date of filing : 06.03.2002 (72)Inventor : HIDAKA KOTA  
NAKAJIMA SHINYA  
MIZUNO OSAMU

(54) VIDEO REPRODUCING METHOD, VIDEO REPRODUCING APPARATUS, AND VIDEO REPRODUCING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a video reproducing method for compressing the contents of a recorded program at an optional summary rate and reproducing the summary of the contents.

SOLUTION: The video reproducing method adopting an audio processing method for deciding an emphasis state on the basis of a feature quantity resulting from analyzing an audio signal by each frame, uses a code table for storing an incident probability in the emphasis state and an incident probability in a still state in cross-reference with each other to obtain the incident probability in the emphasis state corresponding to the feature quantity resulting from analyzing an audio signal by each frame and the incident probability in the still state, calculates the probability of incidence of the emphasis state on the basis of the

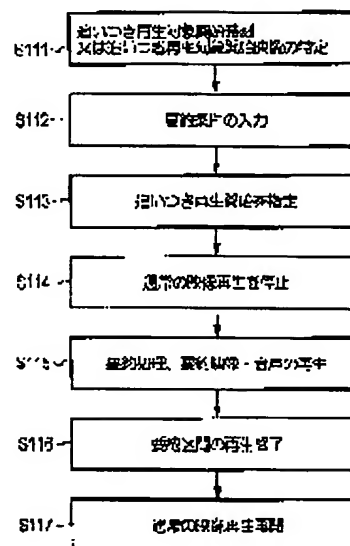


図 1

THIS PAGE BLANK (USPTO)

incidence probability of the audio signal to be summarized recorded on a recording medium and the probability of incidence of the still state on the basis of the incidence probability in the still state, decides an audio signal period wherein a probability ratio of the probability of incidence in the emphasis state to the probability of incidence in the still state is greater than a prescribed coefficient to be a summarized period, calculates the total sum of times in the summarized period, decides the summarized period wherein the summarized time is nearly a prescribed summarized time, and reads the decided summarized period from the recording medium and reproduces it.

---

## LEGAL STATUS

[Date of request for examination] 20.10.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開2003-259311

(P2003-259311A)

(43) 公開日 平成15年9月12日 (2003.9.12)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
H 0 4 N 5/93		H 0 4 N 5/93	Z 5 C 0 5 3
G 1 0 L 11/02		G 1 0 L 3/00	5 1 5 Z 5 D 0 1 5
15/00			5 1 5 E
15/02			5 1 5 B
15/04			5 5 1 G

審査請求 未請求 請求項の数10 O L (全 19 頁) 最終頁に続く

(21) 出願番号 特願2002-60844(P2002-60844)

(22) 出願日 平成14年3月6日 (2002.3.6)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 日▲高▼ 浩太

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(72) 発明者 中島 信弥

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

(74) 代理人 100066153

弁理士 草野 卓 (外1名)

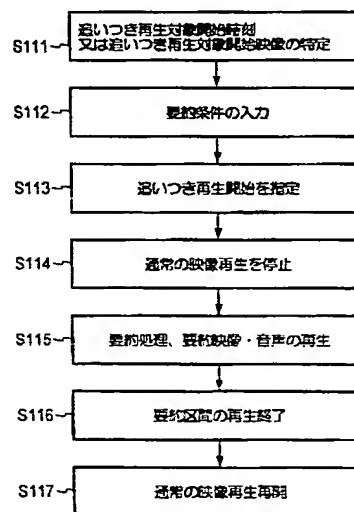
最終頁に続く

(54) 【発明の名称】 映像再生方法、映像再生装置、映像再生プログラム

(57) 【要約】

【課題】 録画されている番組の内容を任意の要約率に圧縮して要約を再生する。

【解決手段】 音声信号をフレーム毎に分析した特徴量に基づき強調状態を判定する音声処理方法において、強調状態での出現確率と平静状態での出現確率とを対応して格納した符号帳を用い、音声信号をフレーム毎に分析した特徴量に対応する強調状態での出現確率と平静状態での出現確率を求め、記録媒体に記録されている要約すべき音声信号の強調状態での出現確率に基づいて強調状態となる確率と、平静状態での出現確率に基づいて平静状態となる確率とを算出し、強調状態となる確率の平静状態となる確率に対する確率比が所定の係数より大きい音声信号区間を要約区間と判定し、要約区間の時間の総和を算出し、要約時間が略所定の要約時間になる要約区間を決定し、決定した要約区間を記録媒体から読み出して再生する映像再生方法。



## 【特許請求の範囲】

【請求項1】 実時間映像信号と音声信号を再生時刻と対応付けて記憶し、要約開始時刻を入力し、要約区間の総延長時間である要約時間又は要約区間の総延長時間の全要約対象区間の比である要約率を入力し、前記要約時間又は要約率で前記要約開始時刻から要約終了時刻として現在までの要約対象区間における音声信号について強調状態と判定された区間を要約区間と判定し、

前記要約区間の音声信号と映像信号を再生することとを特徴とする映像再生方法。

【請求項2】 前記要約区間の音声信号と映像信号の再生終了の時刻を新たな要約区間の終了時刻とし、前記要約区間の再生終了時刻を新たな要約再生開始時刻とする前記要約区間の決定及び当該要約区間の音声信号と映像信号の再生を反復することを特徴とする請求項1記載の映像再生方法。

【請求項3】 前記要約率 $r$  ( $r$ は $0 < r < 1$ となる実数)を $r / (1 + r)$ と調整し、当該調整された要約率をもって要約区間を判定することを特徴とする請求項2記載の映像再生方法。

【請求項4】 少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳を用い、前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確率を算出し、前記平静状態での出現確率に基づいて平静状態となる確率を算出し、前記強調状態となる確率の前記平静状態となる確率に対する確率比を音声信号区間毎に算出し、前記確率比に対応する音声信号区間の時間を降順に累積して要約時間を算出し、前記要約時間の全要約対象区間に対する比である要約率が前記入力された要約率となる音声信号区間を前記要約区間と決定することを特徴とする請求項1乃至3のうち何れか1項に記載の映像再生方法。

【請求項5】 少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳を用い、前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確

率を算出し、

前記平静状態での出現確率に基づいて平静状態となる確率を算出し、

前記強調状態となる確率の前記平静状態となる確率に対する確率比が所定の係数より大きい音声信号区間を要約区間と仮判定し、

要約区間の時間の総和、又は要約率として前記音声信号全区間の時間の前記要約区間の時間の総和に対する比率を算出し、

10 前記要約区間の時間の総和が略所定の要約時間に、又は前記要約率が略所定の要約率となる前記所定の係数を算出して要約区間を決定することを特徴とする請求項1乃至3の何れかに記載の映像再生方法。

【請求項6】 前記音声信号をフレーム毎に無音区間か否か、有声区間か否か判定し、

所定フレーム数以上の無音区間で囲まれ、有声区間を含む部分を音声小段落と判定し、

音声小段落に含まれる有声区間の平均パワーが該音声小段落内の平均パワーの所定の定数倍より小さい音声小段落を末尾とする音声小段落群を音声段落と判定し、

20 前記音声信号区間は音声段落毎に定められたものであり、

前記要約時間を音声段落毎に累積して求めることを特徴とする請求項4又は5の何れかに記載の映像再生方法。

【請求項7】 実時間映像信号と音声信号を再生時刻と対応付けて記憶する記憶手段と、

要約開始時刻を入力する要約開始時刻入力手段と、

要約区間の総延長時間である要約時間又は要約区間の総延長時間の全要約対象区間の比である要約率で定められる要約条件を入力する要約条件入力手段と、

30 前記要約条件に従って、前記要約開始時刻から要約終了時刻として現在までの要約対象区間における音声信号について強調状態と判定された区間を要約区間と判定する要約区間決定手段と、

前記要約区間決定部で決定した要約区間の音声信号と映像信号を再生する再生手段と、

を有することを特徴とする映像再生装置。

【請求項8】 請求項7記載の映像再生装置において、前記要約区間決定手段は、

40 少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳と、

前記符号帳を用いて前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、

前記強調状態での出現確率に基づいて強調状態となる確率を算出する強調状態確率計算部と、

前記符号帳を用いて前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、

50 前記平静状態での出現確率に基づいて平静状態となる確

率を算出する平静状態確率計算部と、  
前記強調状態となる確率の前記平静状態となる確率に対する確率比を音声信号区間毎に算出し、  
前記確率比に対応する音声信号区間の時間を降順に累積して要約時間を算出し、要約区間を仮決定する要約区間仮決定部と、  
前記要約区間の全要約対象区間に対する比が前記要約率を満たす音声信号区間を前記要約区間と決定する要約区間決定部と、  
を有する映像再生装置。

【請求項 9】 請求項 7 記載の映像再生装置において、前記要約区間決定手段は、  
少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳と、  
この符号帳を用いて前記音声符号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率と平静状態での出現確率を求め、  
前記強調状態での出現確率に基づいて強調状態となる確率を算出する強調状態確率計算部と、  
前記平静状態での出現確率に基づいて平静状態となる確率を算出する平静状態確率計算部と、  
前記強調状態となる確率の前記平静状態となる確率に対する確率比が所定の係数より大きい音声信号区間を要約区間と仮判定する要約区間仮判定部と、  
要約区間の時間の総和が略所定の要約時間に、又は前記要約率が略所定の要約率となる前記所定の係数を算出して各チャンネル毎又は各発話者毎の要約区間を決定する要約区間決定部と、  
を有することを特徴とする映像再生装置。

【請求項 10】 コンピュータが解読可能な符号によって記述され、請求項 1 乃至 6 の何れかに記載の映像再生方法を実行させる映像再生プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は記録媒体に記録されている各種の音声付映像を要約して再生する映像再生方法、映像再生装置及び映像再生プログラムに関する。

【0002】

【従来の技術】従来より各種の要約方法が提案されている。その一つとして連続する複数フレームからなる区間動画像を動画全体の各ブロックから抽出し、抽出した各ブロックの区間動画像をつなぎ合わせてダイジェスト画像とする装置があった。例えば、日本国特開平 8-9310 号公報、日本国特開平 3-90968 号公報、日本国特開平 6-165009 号公報などに示されている。また、オーディオセグメントの時間圧縮方法として、ポーズ圧縮の割合を精密に制御し、了解性の高いダイジェストを作成する方法があった。例えば、日本国特開 20

01-154700 公報などに示されている。

【0003】また、テロップや音情報を使って、当該番組映像の特徴となる場面やシーンを抽出してダイジェスト映像とするシステムがあった。例えば、日本国特開 2001-23062 公報などに示されている。

【0004】

【発明が解決しようとする課題】コンテンツを任意の時間で要約、もしくはダイジェストを生成するには、コンテンツを構成する各シーンの優先順位をあらかじめ求めておく必要がある。日本国特開平 8-9310 号公報、日本国特開平 3-90968 号公報、日本国特開平 6-165009 号公報では、ユーザが重要と思うシーンをジョイスティック等のポインティングデバイスや、複数のボタンを用いて入力し、ダイジェスト優先度情報を付与している。利用者にとってダイジェスト生成のための負担が大きい。

【0005】また、日本国特開 2001-154700 公報では、ポーズ圧縮によって、ダイジェストを生成しているが、コンテンツの大半が通例ポーズでない区間で占められている以上、単にポーズを除去するだけでは要約再生時間を元のコンテンツ再生時間の 1/10 以上といった高い圧縮率でコンテンツを圧縮することは非現実的である。また、日本国特開 2000-23062 公報では、ダイジェスト映像生成方法として、音情報の音量値だけを手がかりに特定された要約区間は必ずしも重要な区間とはいえない。何故ならば要点を強調して話す場合、必ずしも音量を大きく話すとは限らないからである。また、テロップ情報を用いる場合、テロップが存在しないコンテンツのダイジェストの生成や、テロップが出現しない区間ではダイジェストを生成することは不可能である。

【0006】また、生放送など実時間映像付音声信号の配信を受け、再生しているとき、離席等により当該番組を視聴できなかった場合に、その前半の部分を録画を続けながら要約して視聴できると、番組前半の筋書きを理解した上で後続する映像を視聴できることが期待される。この発明の目的は記録媒体に格納した映像を任意の時間に圧縮して再生することができる映像再生方法及び映像再生装置、映像再生プログラムを提案しようとするものである。

【0007】

【課題を解決するための手段】この発明では、実時間映像信号と音声信号を再生時刻と対応付けて記憶し、要約開始時刻を入力し、要約区間の総延長時間である要約時間又は要約区間の総延長時間の全要約対象区間の比である要約率を入力し、前記要約時間又は要約率で前記要約開始時刻から要約終了時刻として現在までの要約対象区間における音声信号について強調状態と判定された区間を要約区間と判定し、前記要約区間の音声信号と映像信号を再生する映像再生方法を提案する。

【0008】この発明では更に、前記要約区間の音声信号と映像信号の再生終了の時刻を新たな要約区間の終了時刻とし、前記要約区間の再生終了時刻を新たな要約区間再生開始時刻とする前記要約区間の決定及び当該要約区間の音声信号と映像信号の再生を反復する映像再生方法を提案する。この発明では更に、前記要約率 $r$  ( $r$ は $0 < r < 1$ となる実数)を $r / (1 + r)$ と調整し、当該調整された要約率をもって要約区間を判定する映像再生方法を提案する。この発明では更に、少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳を用い、前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確率を算出し、前記平静状態での出現確率に基づいて平静状態となる確率を算出し、前記強調状態となる確率の前記平静状態となる確率に対する確率比を音声信号区間毎に算出し、前記確率比に対応する音声信号区間の時間を降順に累積して要約時間を算出し、前記要約時間の全要約対象区間に対する比である要約率が前記入力された要約率となる音声信号区間を前記要約区間と決定する映像再生方法を提案する。

【0009】この発明では更に、少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳を用い、前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確率を算出し、前記平静状態での出現確率に基づいて平静状態となる確率を算出し、前記強調状態となる確率の前記平静状態となる確率に対する確率比が所定の係数より大きい音声信号区間を要約区間と仮判定し、要約区間の時間の総和、又は要約率として前記音声信号全区間の時間の前記要約区間の時間の総和に対する比率を算出し、前記要約区間の時間の総和が略所定の要約時間に、又は前記要約率が略所定の要約率となる前記所定の係数を算出して要約区間を決定する映像再生方法を提案する。

【0010】この発明では更に、前記音声信号をフレーム毎に無音区間か否か、有音区間か否か判定し、所定フレーム数以上の無音区間で囲まれ、有音区間を含む部分を音声小段落と判定し、音声小段落に含まれる有音区間の平均パワーが該音声小段落内の平均パワーの所定の定数倍より小さい音声小段落を末尾とする音声小段落群を音声段落と判定し、前記音声信号区間は音声段落毎に定

められたものであり、前記要約時間を音声段落毎に累積して求める映像再生方法を提案する。

【0011】この発明では更に、実時間映像信号と音声信号を再生時刻と対応付けて記憶する記憶手段と、要約開始時刻を入力する要約開始時刻入力手段と、要約区間の総延長時間である要約時間又は要約区間の総延長時間の全要約対象区間の比である要約率で定められる要約条件を入力する要約条件入力手段と、前記要約条件に従って、前記要約開始時刻から要約終了時刻として現在までの要約対象区間における音声信号について強調状態と判定された区間を要約区間と判定する要約区間決定手段と、前記要約区間決定部で決定した要約区間の音声信号と映像信号を再生する再生手段とを有する映像再生装置を提案する。

【0012】この発明では更に、前記映像再生装置において、前記要約区間決定手段は、少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳と、前記符号帳を用いて前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確率を算出する強調状態確率計算部と、前記符号帳を用いて前記音声信号をフレーム毎に分析した前記特徴量に対応する平静状態での出現確率を求め、前記平静状態での出現確率に基づいて平静状態となる確率を算出する平静状態確率計算部と、前記強調状態となる確率の前記平静状態となる確率に対する確率比を音声信号区間毎に算出し、前記確率比に対応する音声信号区間の時間を降順に累積して要約時間を算出し、要約区間を仮決定する要約区間仮決定部と、前記要約区間の全要約対象区間に対する比が前記要約率を満たす音声信号区間を前記要約区間と決定する要約区間決定部とを有する映像再生装置を提案する。

【0013】この発明では更に、前記映像再生装置において、前記要約区間決定手段は、少なくとも基本周波数又はピッチ周期、パワー、動的特徴量の時間変化特性、又はこれらのフレーム間差分を含む特徴量と強調状態での出現確率と平静状態での出現確率を対応して格納した符号帳と、この符号帳を用いて前記音声信号をフレーム毎に分析した前記特徴量に対応する強調状態での出現確率と平静状態での出現確率を求め、前記強調状態での出現確率に基づいて強調状態となる確率を算出する強調状態確率計算部と、前記平静状態での出現確率に基づいて平静状態となる確率を算出する平静状態確率計算部と、前記強調状態となる確率の前記平静状態となる確率に対する確率比が所定の係数より大きい音声信号区間を要約区間と仮判定する要約区間仮判定部と、要約区間の時間の総和が略所定の要約時間に、又は前記要約率が略所定の要約率となる前記所定の係数を算出して各チャンネル毎



又は各発話者毎の要約区間を決定する要約区間決定部とを有する映像再生装置。

【0014】この発明では更に、コンピュータが解読可能な符号によって記述され、前記記載の映像再生方法の何れかを実行させる映像再生プログラムを提案する。

#### 作用

この発明の映像再生方法によれば記録媒体に記録されている音声の強調状態となる確率が高い音声区間を要約区間として抽出するから、コンテンツの内容で重要な部分を抜き出し、重要な部分をつなぎ合せて要約音声及び要約された映像情報を得ることができる。この結果、要約時間を短時間に圧縮したとしても、そのコンテンツの内容をよく理解することができる。

【0015】また、記録媒体として短時間に多量のデータを書き込み及び読み出すことができる記録媒体を用いることから、録画を続けながら、他のコンテンツ又は録画中のコンテンツの前半部分を読み出すことができる。このために、録画を続けながら、その録画中の番組の録画部分を要約し、要約情報を再生することができる。この結果、録画中の映像と、要約を伝える映像を例えば親画面と子画面と異なる表示手段にそれぞれに表示することにより、現在の放送内容と過去の放送内容の双方を視ることができ、要約情報の再生が終了した時点では番組の前半の部を理解した状態で続きを視聴することができる利点を得られる。

【0016】この発明の特徴とする点は、コンテンツ要約再生時にユーザからの要求に従って、どのような要約率（圧縮率）にでもコンテンツを要約することができる要約方法を用いる点にある。この特徴とする要約方法は、先願である特願2001-241278で本出願人が提案した、任意の音声小段落の発話状態を判定し、強調状態となる確率が平静状態となる確率よりも大きければ、その音声小段落を強調状態にあると判定し、その音声小段落を含む音声段落を要約区間として抽出する音声強調状態判定方法及び音声要約方法を利用して実現することができる。

【0017】

【発明の実施の形態】ここで、この発明で用いられる音声小段落抽出方法、音声段落抽出方法、各音声小段落毎に強調状態となる確率及び平静状態となる確率を求める方法について、説明する。図5に先に提案した音声要約方法の実施形態の基本手順を示す。ステップS1で入力音声信号を分析して音声特徴量を求める。ステップS2で、入力音声信号の音声小段落と、複数の音声小段落から構成される音声段落を抽出する。ステップS3で各音声小段落を構成するフレームが平静状態か、強調状態か発話状態を判定する。この判定に基づきステップS4で要約音声を作成し、要約音声を得る。

【0018】以下に、自然な話し言葉や会話音声、要約に適用する場合の実施例を述べる。音声特徴量は、ス

ベクトル情報等に比べて、雑音環境下でも安定して得られ、かつ話者に依存し難いものを用いる。入力音声信号から音声特徴量として基本周波数（ $f_0$ ）、パワー

（ $p$ ）、音声の動的特徴量の時間変化特性（ $d$ ）、ポーズ時間長（無音区間）（ $p_s$ ）を抽出する。これらの音声特徴量の抽出法は、例えば、「音響・音響工学」（古井貞照、近代科学社、1998）、「音声符号化」（守谷健弘、電子情報通信学会、1998）、「デジタル音声処理」（古井貞照、東海大学出版会、1985）、

10 「複合正弦波モデルに基づく音声分析アルゴリズムに関する研究」（嵯峨山茂樹、博士論文、1998）などに述べられている。音声の動的特徴量の時間変化は発話速度の尺度となるパラメータであり特許第2976998号に記載のものを用いてもよい。即ち、動的变化量としてスペクトル包絡を反映するLPCスペクトラム係数の時間変化特性を求め、その時間変化をもとに発話速度係数が求められるものである。より具体的にはフレーム毎にLPCスペクトラム係数 $C_1(t)$ 、 $\dots$ 、 $C_k(t)$ を抽出して次式のような動的特徴量 $d$ （ダイナミックメジャー）を求める。 $d(t) = \sum_{i=1}^k [\sum_{f=t-f_0t+f_0} [f \times C_i(t)] / (\sum_{f=t-f_0t+f_0} f^2)]$ ここで、 $f_0$ は前後の音声区間フレーム数（必ずしも整数個のフレームでなくとも一定の時間区間でもよい）、 $k$ はLPCスペクトラムの次数、 $i=1, 2, \dots, k$ である。発話速度の係数として動的特徴量の変化の極大点の単位時間当たりの個数、もしくは単位時間当たりの変化率が用いられる。

【0019】実施例では例えば100msを1フレームとし、シフトを50msとする。1フレーム毎の平均の基本周波数を求める（ $f_0'$ ）。パワーについても同様に1フレーム毎の平均パワー（ $p'$ ）を求める。更に現フレームの $f_0'$ と $\pm i$ フレーム前後の $f_0'$ との差分をとり、 $\pm \Delta f_0' i$ （ $\Delta$ 成分）とする。パワーについても同様に現フレームの $p'$ と $\pm i$ フレーム前後の $p'$ との差分 $\pm \Delta p' i$ （ $\Delta$ 成分）を求める。 $f_0'$ 、 $\pm \Delta f_0' i$ 、 $p'$ 、 $\pm \Delta p' i$ を規格化する。この規格は例えば $f_0'$ 、 $\pm \Delta f_0' i$ をそれぞれ、音声波形全体の平均基本周波数で割り規格化する。これら規格化された値を $f_0''$ 、 $\pm f_0'' i$ と表す。 $p'$ 、 $\pm \Delta p' i$ についても同様に、発話状態判定の対象とする音声波形全体の平均パワーで割り、規格化する。規格化するにあたり、後述する音声小段落、音声段落ごとの平均パワーで割ってもよい。これら規格化された値を $p''$ 、 $\pm \Delta p'' i$ と表す。 $i$ の値は例えば $i=4$ とする。現フレームの前後 $\pm T_1$ msの、ダイナミックメジャーのピーク本数、即ち動的特徴量の変化の極大点の個数を数える（ $d_p$ ）。これと、現フレームの開始時刻の、 $T_2$ ms前の時刻を区間に含むフレームの $d_p$ との $\Delta$ 成分（ $-\Delta d_p$ ）を求める。前記 $\pm T_1$ msの $d_p$ 数と、現フレームの終了時刻の、 $T_3$ ms後の時刻を区間に含むフレーム

の $d_p$ との $\Delta$ 成分(+ $\Delta d_p$ )を求める。これら、 $T_1$ 、 $T_2$ 、 $T_3$ の値は例えば $T_1 = T_2 = T_3 = 450$  msとする。フレームの前後の無音区間の時間長を+ $n$  sとする。ステップS1ではこれら音声特徴パラメータの各値をフレーム毎に抽出する。

【0020】ステップS2における入力音声の音声小段落と、音声段落を抽出する方法の例を図6に示す。ここで音声小段落を発話状態判定を行う単位とする。ステップS201で、入力音声信号の無音区間と有声区間を抽出する。無音区間は例えばフレーム毎のパワーが所定の

10

20

パワー値以下であれば無音区間と判定し、有声区間は、例えばフレーム毎の相関関数が所定の相関関数値以上であれば有声区間と判定する。有声/無声の決定は、周期性/非周期性の特徴と同一視することにより、自己相関関数や変形相関関数のピーク値で行うことが多い。入力信号の短時間スペクトルからスペクトル包絡を除去した予測残差の自己相関関数が変形相関関数であり、変形相関関数のピークが所定の閾値より大きいかなにかによって有声/無声の判定を行い、又そのピークを与える遅延時間によってピッチ周期 $1/f_0$ （基本周波数 $f_0$ ）の抽出を行う。これらの区間の抽出法の詳細は、例えば、「デジタル音声処理」（古井貞照、東海大学出版会、1985）などに述べられている。ここでは音声信号から各音声特徴量をフレーム毎に分析することについて述\*

$$\sum p_i / (\alpha + 1) < \beta P_j$$

を満たす時、音声小段落 $j$ が音声段落 $k$ の末尾音声小段落であるとする。ただし、 $\Sigma$ は $i = n - \alpha$ から $n$ までである。式(1)の $\alpha$ 、 $\beta$ は定数であり、これら进行操作して、音声段落を抽出する。実施例では、 $\alpha$ は3、 $\beta$ は0.8とした。このようにして末尾音声小段落を区切りとして隣接する末尾音声小段落間の音声小段落群を音声段落と判定できる。

【0022】図5中のステップS3における音声小段落発話状態判定方法の例を図8に示す。ステップS301で、入力音声小段落の音声特徴量をベクトル量子化する。このために、あらかじめ少なくとも2つの量子化音声特徴量（コード）が格納された符号帳（コードブック）を作成しておく。ここでコードブックに蓄えられた音声特徴量と入力音声もしくは既に分析して得られた音声の音声特徴量との照合をとり、コードブックの中から音声特徴量間の歪（距離）を最小にする量子化音声特徴量を特定することが常套である。図9に、このコードブックの作成法の例を示す。多数の学習用音声を被験者が聴取し、発話状態が平静状態であるものと、強調状態であるものをラベリングする（S501）。

【0023】例えば、被験者が発話の中で強調状態とする理由として、

(a) 声が大きく、名詞や接続詞を伸ばすように発話する

(b) 話し始めを伸ばして話題変更を主張、意見を集約

\* べたが、既に符号化等により分析された係数もしくは符号に対応する特徴量を符号化に用いる符号帳から読み出して用いてもよい。

【0021】ステップS202で、有声区間を囲む無音区間の時間がそれぞれ $t$ 秒以上になるとき、その無音区間で囲まれた有声区間を含む部分を音声小段落とする。この $t$ は例えば $t = 400$  msとする。ステップS203で、この音声小段落内の好ましくは後半部の、有声区間の平均パワーと、その音声小段落の平均のパワーの値BAの定数 $\beta$ 倍とを比較し、前者の方が小さい場合はその音声小段落を末尾音声小段落とし、直前の末尾音声小段落後の音声小段落から現に検出した末尾音声小段落までを音声段落として決定する。図7に、有声区間、音声小段落、音声段落を模式的に示す。音声小段落を前記の、有声区間を囲む無音区間の時間が $t$ 秒の条件で、抽出する。図7では、音声小段落 $j-1$ 、 $j$ 、 $j+1$ について示している。ここで音声小段落 $j$ は、 $n$ 個の有声区間から構成され、平均パワーを $P_j$ とする。有声区間の典型的な例として、音声小段落 $j$ に含まれる、有声区間 $v$ の平均パワーは $p_v$ である。音声段落 $k$ は、音声小段落 $j$ と音声小段落を構成する後半部分の有声区間のパワーから抽出する。 $i = n - \alpha$ から $n$ までの有声区間の平均パワー $p_i$ の平均が音声小段落 $j$ の平均パワー $P_j$ より小さいとき、即ち、

式(1)

するように声を大きくする

(c) 声を大きく高くして重要な名詞等を強調する時

(d) 高音であるが声はそれほど大きくない

(e) 苦笑いしながら、焦りから本音をこまかすような時

(f) 周囲に同意を求める、あるいは問いかけるように、語尾が高音になるとき

(g) ゆっくりと力強く、念を押すように、語尾の声が大きくなる時

(h) 声が大きく高く、割り込んで発話するという主張、相手より大きな声で

(i) 大きな声では憚られるような本音や秘密を発言する場合や、普段、声の大きい人にとっての重要なことを発話するような時（例えば声が小さくボソボソ、ヒソヒソという口調）を挙げた。この例では、平静状態とは、前記の(a)～(i)のいずれでもなく、発話が平静であると被験者が感じたものとした。

【0024】尚、上述では強調状態と判定する対象を発話であるものとして説明したが、音楽でも強調状態を特定することができる。ここでは音声付の楽曲において、音声から強調状態を特定しようとした場合に、強調と感

(a) 声が大きく、かつ声が高い

(b) 声が力強い

(c) 声が高く、かつアクセントが強い

50

- (d) 声が高く、声質が変化する
- (e) 声を伸長させ、かつ声大きい
- (f) 声が大きく、かつ、声が高く、アクセントが強い
- (g) 声が大きく、かつ、声が高く、叫んでいる
- (h) 声が高く、アクセントが変化する
- (i) 声を伸長させ、かつ、声大きく、語尾が高い
- (j) 声が高く、かつ、声を伸長させる
- (k) 声を伸長させ、かつ、叫び、声が高い
- (l) 語尾上がり力強い
- (m) ゆっくり強い
- (n) 曲調が不規則
- (o) 曲調が不規則、かつ、声が高い

また、音声を含まない楽器演奏のみの楽曲でも強調状態を特定することができる。その強調と感ずる理由として、

- (a) 強調部分全体のパワー増大
- (b) 音の高低差が大きい
- (c) パワーが増大する
- (d) 楽器の数が変化する
- (e) 曲調、テンポが変化する

等である。

【0025】これらを基にコードブックを作成しておくことにより、発話に限らず音楽の要約も行うことができることになる。平静状態と強調状態の各ラベル区間について、図5中のステップS1と同様に、音声特徴量を抽出し(S502)、パラメータを選択する(S503)。平静状態と強調状態のラベル区間の、前記パラメータを用いて、LBGアルゴリズムでコードブックを作成する(S504)。LBGアルゴリズムについては、例えば、(Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol. Com-28, pp. 84-95, 1980)がある。コードブックサイズは2のn乗個に可変である。このコードブック作成は音声小段落で又はこれより長い適当な区間毎あるいは学習音声全体の音声特徴量で規格化した音声特徴量を用いることが好ましい。

【0026】図8中のステップS301で、このコードブックを用いて、入力音声小段落の音声特徴量を、各音\*

$$P_{\alpha}(e) = P_{\text{emp}}(C_n | C_{n-N+1} \dots C_{n-1})$$

$$P_{\alpha}(n) = P_{\text{nrm}}(C_n | C_{n-N+1} \dots C_{n-1})$$

とする。つまり $C_n$ より $N-1$ 個の過去のコード列 $C_{n-N+1} \dots C_{n-1}$ の次に $C_n$ が強調状態として得られる確率を $P_{\alpha}(e)$ とし、同様にN-gramの確率値をより低次のM-gram( $N \geq M$ )の確率値と線形に補間する線形補間法を適応することが好ましい。例えば $C_n$ より $N-1$ 個の過去のコード列 $C_{n-N+1} \dots C_{n-1}$ の次に $C_n$ が平静状態として得られる確率を $P_{\alpha}(n)$ とする。このような $P_{\alpha}(e)$ 、 $P_{\alpha}(n)$ の条

\* 声特徴量について規格化し、その規格化された音声特徴量をフレーム毎に照合もしくはベクトル量子化し、フレーム毎にコード(量子化された音声特徴量)を得る。この際の入力音声信号より抽出する音声特徴量は前記のコードブック作成に用いたパラメータと同じである。強調状態が含まれる音声小段落を特定するために、音声小段落でのコードを用いて、発話状態の尤度(らしさ)を、平静状態と強調状態について求める。このために、あらかじめ、任意のコード(量子化音声特徴量)の出現確率を、平静状態の場合と、強調状態の場合について求めておき、この出現確率とそのコードとを組としてコードブックに格納しておく、以下にこの出現確率の求め方の例を述べる。前記のコードブック作成に用いた学習音声中のラベルが与えられた1つの区間(ラベル区間)の音声特徴量のコード(フレーム毎に得られる)が、時系列で $C_i, C_j, C_k, \dots, C_n$ であるとき、ラベル区間 $\alpha$ が強調状態となる確率を $P_{\alpha}(e)$ 、平静状態となる確率を $P_{\alpha}(n)$ とし、

$$P_{\alpha}(e) = P_{\text{emp}}(C_i) P_{\text{emp}}(C_j | C_i)$$

$$\dots P_{\text{emp}}(C_n | C_i \dots C_{n-1}) = P_{\text{emp}}(C_i) \Pi P_{\text{emp}}(C_x | C_i \dots C_{x-1})$$

$$P_{\alpha}(n) = P_{\text{nrm}}(C_i) P_{\text{nrm}}(C_j | C_i) \dots P_{\text{nrm}}(C_n | C_i \dots C_{n-1}) = P_{\text{emp}}(C_i) \Pi P_{\text{nrm}}(C_x | C_i \dots C_{x-1})$$

となる。ただし、 $P_{\text{emp}}(C_x | C_i \dots C_{x-1})$ はコード列 $C_i \dots C_{x-1}$ の次に $C_x$ が強調状態となる条件付確率、 $P_{\text{nrm}}(C_x | C_i \dots C_{x-1})$ は同様に $C_i \dots C_{x-1}$ に対し $C_x$ が平静状態となる確率である。ただし、 $\Pi$ は $x = i+1$ から $n$ までの積である。また $P_{\text{emp}}(C_i)$ は学習音声についてフレームで量子化し、これらコード中の $C_i$ が強調状態とラベリングされた部分に存在した個数を計数し、その計数値を全学習音声の全コード数(フレーム数)で割り算した値であり、 $P_{\text{nrm}}(C_i)$ は $C_i$ が平静状態とラベリングされた部分に存在した個数を全コード数で割り算した値である。

【0027】このラベル区間 $\alpha$ の各状態確率を簡単にするために、この例ではN-gramモデル( $N < n$ )を用いて、

条件確率をラベリングされた学習音声の量子化コード列から全てを求めるが、入力音声信号の音声特徴量の量子化したコード列と対応するものが学習音声から得られていない場合もある。そのため、高次(即ちコード列の長い)の条件付確率を単独出現確率とより低次の条件付出現確率とを補間して求める。例えば $N=3$ のtrigram、 $N=2$ のbigram、 $N=1$ のunigramを用いて線形補間法を施す。N-gram、線形補間

法、trigramについては、例えば、「音声言語処理」(北 研二、中村 哲、永田昌明、森北出版、1996、29頁)などに述べられている。即ち、

$N=3$  (trigram) :  $P_{\text{emp}}(C_n | C_{n-2} C_{n-1})$ 、 $P_{\text{nr}}(C_n | C_{n-2} C_{n-1})$

$N=2$  (bigram) :  $P_{\text{emp}}(C_n | C_{n-1})$ 、 $P_{\text{nr}}(C_n | C_{n-1})$

$$P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) = \lambda_{\text{emp}1} P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) + \lambda_{\text{emp}2} P_{\text{emp}}(C_n | C_{n-1}) + \lambda_{\text{emp}3} P_{\text{emp}}(C_n)$$

式(2)

$$P_{\text{nr}}(C_n | C_{n-2} C_{n-1}) = \lambda_{\text{nr}1} P_{\text{nr}}(C_n | C_{n-2} C_{n-1}) + \lambda_{\text{nr}2} P_{\text{nr}}(C_n | C_{n-1}) + \lambda_{\text{nr}3} P_{\text{nr}}(C_n)$$

式(3)

Trigramの学習データをNとしたとき、すなわち、コードが時系列で $C_1, C_2, \dots, C_N$ が得られたとき、 $\lambda_{\text{emp}1}$ 、 $\lambda_{\text{emp}2}$ 、 $\lambda_{\text{emp}3}$ の再推定式は前出の参考文献「音声言語処理」より次のようになる。

$\lambda_{\text{emp}1} = 1 / N \sum (\lambda_{\text{emp}1} P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) / (\lambda_{\text{emp}1} P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) + \lambda_{\text{emp}2} P_{\text{emp}}(C_n | C_{n-1}) + \lambda_{\text{emp}3} P_{\text{emp}}(C_n))$

$\lambda_{\text{emp}2} = 1 / N \sum (\lambda_{\text{emp}2} P_{\text{emp}}(C_n | C_{n-1}) / (\lambda_{\text{emp}1} P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) + \lambda_{\text{emp}2} P_{\text{emp}}(C_n | C_{n-1}) + \lambda_{\text{emp}3} P_{\text{emp}}(C_n))$

$$P_{\alpha}(e) = P_{\text{emp}}(C_{i3} | C_{i1} C_{i2}) \dots P_{\text{emp}}(C_{iN_{\alpha}} | C_{i(N_{\alpha}-1)} C_{i(N_{\alpha}-2)})$$

式(4)

$$P_{\alpha}(n) = P_{\text{nr}}(C_{i3} | C_{i1} C_{i2}) \dots P_{\text{nr}}(C_{iN_{\alpha}} | C_{i(N_{\alpha}-1)} C_{i(N_{\alpha}-2)})$$

式(5)

となる。この計算ができるように前記のtrigram、unigram、bigramを任意のコードについて求めてコードブックに格納しておく。つまりコードブックには各コードの音声特徴量とその強調状態での出現確率とこの例では平静状態での出現確率との組が格納され、その強調状態での出現確率は、その音声特徴量が過去のフレームでの音声特徴量と無関係に強調状態で出現する確率(unigram: 単独出現確率と記す)のみ、又はこれと、過去のフレームでの音声特徴量から現在のフレームの音声特徴量に至るフレーム単位の音声特徴量列毎に、その音声特徴量が強調状態で出現する条件付確率との組合せの何れかであり、平静状態での出現確率も同様に、その音声特徴量が過去のフレームでの音声特徴量と無関係に平静状態で出現する確率(unigram: 単独出現確率と記す)のみ、又はこれと、過去のフレームでの音声特徴量から現在のフレームの音声特徴量に至るフレーム単位の音声特徴量列毎にその音声特徴量が平静状態で出現する条件付確率と組合せの何れかである。

【0029】例えば図10に示すようにコードブックには各コード $C_1, C_2, \dots$ 毎にその音声特徴量と、その

\*  $N=1$  (unigram) :  $P_{\text{emp}}(C_n)$ 、 $P_{\text{nr}}(C_n)$

であり、これら3つの強調状態での $C_n$ の出現確率、また3つの平静状態での $C_n$ の出現確率をそれぞれ用いて次式により、 $P_{\text{emp}}(C_n | C_{n-2} C_{n-1})$ 、 $P_{\text{nr}}(C_n | C_{n-2} C_{n-1})$ を計算することにする。

※  $3 P_{\text{emp}}(C_n)$  ) )

$\lambda_{\text{emp}3} = 1 / N \sum (\lambda_{\text{emp}3} P_{\text{emp}}(C_n) / (\lambda_{\text{emp}1} P_{\text{emp}}(C_n | C_{n-2} C_{n-1}) + \lambda_{\text{emp}2} P_{\text{emp}}(C_n | C_{n-1}) + \lambda_{\text{emp}3} P_{\text{emp}}(C_n))$

ただし、 $\Sigma$ は $n=1$ から $N$ までの和である。以下同様にして $\lambda_{\text{nr}1}$ 、 $\lambda_{\text{nr}2}$ 、 $\lambda_{\text{nr}3}$ も求められる。

【0028】この例では、ラベル区間 $\alpha$ がフレーム数 $N_{\alpha}$ で得たコードが $C_{i1}, C_{i2}, \dots, C_{iN_{\alpha}}$ のとき、このラベル区間 $\alpha$ が強調状態となる確率 $P_{\alpha}(e)$ 、平静状態となる確率 $P_{\alpha}(n)$ は、

30 単独出現確率が強調状態、平静状態について、また条件付確率が強調状態、平静状態についてそれぞれ組として格納されている。図8中のステップS302では、入力音声小段落の全フレームのコードについてのそのコードブックに格納されている前記確率から、発話状態の尤度を、平静状態と強調状態について求める。図11に実施例の模式図を示す。時刻 $t$ から始まる音声小段落のうち、第4フレームまでを①～④で示している。前記のように、ここでは、フレーム長は100ms、フレームシフトを50msとフレーム長の方を長くした。①フレーム番号 $f$ 、時刻 $t \sim t+100$ でコード $C_i$ が、②フレーム番号 $f+1$ 、時刻 $t+50 \sim t+150$ でコード $C_j$ が、③フレーム番号 $f+2$ 、時刻 $t+100 \sim t+200$ でコード $C_k$ が、④フレーム番号 $f+3$ 、時刻 $t+150 \sim t+250$ でコード $C_l$ が得られ、つまりフレーム順にコードが $C_i, C_j, C_k, C_l$ であるとき、フレーム番号 $f+2$ 以上のフレームでtrigramが計算できる。音声小段落 $s$ が強調状態となる確率を $P_s(e)$ 、平静状態となる確率を $P_s(n)$ とすると第4フレームまでの確率はそれぞれ、

$$P_s(e) = P_{emp}(C_k | C_i C_j) P_{emp}(C_l | C_j C_k) \quad \text{式(6)}$$

$$P_s(n) = P_{nrm}(C_k | C_i C_j) P_{nrm}(C_l | C_j C_k) \quad \text{式(7)}$$

となる。ただし、この例では、コードブックから  $C_k$ 、 $C_l$  の強調状態及び平静状態の各単独出現確率を求め、また  $C_j$  の次に  $C_k$  が強調状態及び平静状態で各出現する条件付確率、更に  $C_k$  が  $C_i$ 、 $C_j$  の次に、 $C_l$  が  $C_k$

\*  $j$ 、 $C_k$  の次にそれぞれ強調状態及び平静状態でそれぞれ出現する条件付確率をコードブックから求めると、以下のようになる。

$$P_{emp}(C_k | C_i C_j) = \lambda_{emp1} P_{emp}(C_k | C_i C_j) + \lambda_{emp2} P_{emp}(C_k | C_j) + \lambda_{emp3} P_{emp}(C_k) \quad \text{式(8)}$$

$$P_{emp}(C_l | C_j C_k) = \lambda_{emp1} P_{emp}(C_l | C_j C_k) + \lambda_{emp2} P_{emp}(C_l | C_k) + \lambda_{emp3} P_{emp}(C_l) \quad \text{式(9)}$$

$$P_{nrm}(C_k | C_i C_j) = \lambda_{nrm1} P_{nrm}(C_k | C_i C_j) + \lambda_{nrm2} P_{nrm}(C_k | C_j) + \lambda_{nrm3} P_{nrm}(C_k) \quad \text{式(10)}$$

$$P_{nrm}(C_l | C_j C_k) = \lambda_{nrm1} P_{nrm}(C_l | C_j C_k) + \lambda_{nrm2} P_{nrm}(C_l | C_k) + \lambda_{nrm3} P_{nrm}(C_l) \quad \text{式(11)}$$

上記(8)～(11)式を用いて(6)式と(7)式で示される第4フレームまでの強調状態となる確率  $P_s$

( $e$ )と、平静状態となる確率  $P_s(n)$  が求まる。ここで、 $P_{emp}(C_k | C_i C_j)$ 、 $P_{nrm}(C_k | C_i C_j)$  はフレーム番号  $f+2$  において計算できる。  
【0030】この例では、音声小段落  $s$  がフレーム数  $N_s$  で得たコードが  $C_{i1}$ 、 $C_{i2}$ 、…、 $C_{iN_s}$  のとき、この音声小段落  $s$  が強調状態になる確率  $P_s(e)$  と平静状態になる確率  $P_s(n)$  を次式により計算する。

$$P_s(e) = P_{emp}(C_{i3} | C_{i1} C_{i2}) \cdots P_{emp}(C_{iN_s} | C_{i(N_s-1)} C_{i(N_s-2)})$$

$$P_s(n) = P_{nrm}(C_{i3} | C_{i1} C_{i2}) \cdots P_{nrm}(C_{iN_s} | C_{i(N_s-1)} C_{i(N_s-2)})$$

この例ではこれらの確率が、 $P_s(e) > P_s(n)$  であれば、その音声小段落  $s$  は強調状態、 $P_s(n) > P_s(e)$  であれば平静状態とする。

【0031】図12は以上説明した音声小段落抽出方法、音声段落抽出方法、各音声小段落毎に強調状態となる確率及び平静状態となる確率を求める方法を用いた音声強調状態判定装置及び音声要約装置の実施形態を示す。入力部11に音声強調状態が判定されるべき、又は音声の要約区間を決定されるべき入力音声(入力音声信号)が入力される。入力部11には必要に応じて入力音声信号をデジタル信号に変換する機能も含まれる。デジタル化された音声信号は必要に応じて記憶部12に格納される。音声特徴量抽出部13で前述した音声特徴量がフレーム毎に抽出される。抽出した音声特徴量は必要に応じて、音声特徴量の平均値で規格化され、量子化部14で各フレームの音声特徴量がコードブック15を参照して量子化され、量子化された音声特徴量は強調確率計算部16と平静確率計算部17に送り込まれる。コードブック15は例えば図10に示したようなものである。

【0032】強調確率計算部16によりその量子化され

た音声特徴量の強調状態での出現確率が、コードブック15に格納されている対応する確率を用いて、例えば式(8)又は(9)により計算される。同様に平静確率計算部17により、前記量子化された音声特徴量の平静状態での出現確率がコードブック15に格納されている対応する確率を用いて、例えば式(10)又は(11)により計算される。強調確率計算部16及び平静確率計算部17で各フレーム毎に算出された強調状態での出現率と平静状態での出現率及び各フレームの音声特徴量は各フレームに付与したフレーム番号と共に記憶部12に格納する。

【0033】これら各部の制御は制御部19の制御のもとに順次行われる。音声要約装置の実施形態は、図12中に実線ブロックに対し、破線ブロックが付加される。つまり記憶部12に格納されている各フレームの音声特徴量が無音区間判定部21と有音区間判定部22に送り込まれ、無音区間判定部21により各フレーム毎に無音区間か否かが判定され、また有音区間判定部22により各フレーム毎に有音区間か否かが判定される。これらの無音区間判定結果と有音区間判定結果が音声小段落判定部23に入力される。音声小段落判定部23はこれら無音区間判定、有音区間判定に基づき、先の方法の実施形態で説明したように所定フレーム数を連続する無音区間に囲まれた有音区間を含む部分が音声小段落と判定する。音声小段落判定部23の判定結果は記憶部12に書き込まれ、記憶部12に格納されている音声データ列に付記され、無音区間で囲まれたフレーム群に音声小段落番号列を付与する。これと共に音声小段落判定部23の判定結果は末尾音声小段落判定部24に入力される。

【0034】末尾音声小段落判定部24では、例えば図7を参照して説明した手法により末尾音声小段落が検出され、末尾音声小段落判定結果が音声段落判定部25に入力され、音声段落判定部25により2つの末尾音声小段落間の複数の音声小段落を含む部分を音声段落と判定

する。この音声段落判定結果も記憶部12に書き込まれ、記憶部12に記憶している音声小段落番号列に音声段落列番号を付与する。音声要約装置として動作する場合、強調確率計算部16及び平静確率計算部17では記憶部12から各音声小段落を構成する各フレームの強調確率と平静確率を読み出し、各音声小段落毎の確率が例えば式(8)及び式(10)により計算される。強調状態判定部18ではこの音声小段落毎の確率計算値を比較して、その音声小段落が強調状態か否かを判定し、要約区間取出し部26では音声段落中の1つの音声小段落でも強調状態と判定されたものがあればその音声小段落を含む音声段落を取り出す。各部の制御は制御部19により行われる。

【0035】以上により音声で構成される音声波形を音声小段落及び音声段落に分離する方法及び各音声小段落毎に強調状態となる確率及び平静状態となる確率を算出できることが理解できよう。次に上述した各方法を利用して要約率を自由に設定し、変更することができる音声処理方法、音声処理装置に関わる実施の形態を説明する。図13にその音声処理方法の実施の形態の基本手順を示す。この実施例ではステップS11で音声強調確率算出処理を実行し、音声小段落の強調確率及び平静確率を求める。

【0036】ステップS12では要約条件入力ステップS12を実行する。この要約条件入力ステップS12では例えば利用者に要約時間又は要約率又は圧縮率の入力を促す情報を提供し、要約時間又は要約率又は要約率又は圧縮率を入力させる。尚、予め設定された複数の要約時間又は要約率、圧縮率の中から一つを選択する入力方法を探ることもできる。ステップS13では抽出条件の変更を繰り返す動作を実行し、ステップS12の要約条件入力ステップS12で入力された要約時間又は要約率、圧縮率を満たす抽出条件を決定する。

【0037】ステップS14で要約抽出ステップを実行する。この要約抽出ステップS14では抽出条件変更ステップS13で決定した抽出条件を用いて採用すべき音声段落を決定し、この採用すべき音声段落の総延長時間を計算する。ステップ15では要約再生処理を実行し、要約抽出ステップS14で抽出した音声段落列を再生する。図14は図13に示した音声強調確率算出ステップの詳細を示す。ステップS101で要約対象とする音声波形列を音声小段落に分離する。ステップS102ではステップS101で分離した音声小段落列から音声段落を抽出する。音声段落とは図7で説明したように、1つ以上の音声小段落で構成され、意味を理解できる単位である。

【0038】ステップS103及びステップS104でステップS101で抽出した音声小段落毎に図10で説明したコードブックと前記した式(8)、(10)等を利用して各音声小段落が強調状態となる確率(以下強調

確率と称す)  $P_s(e)$  と、平静状態となる確率(以下平静確率と称す)  $P_s(n)$  とを求める。ステップS105ではステップS103及びS104において各音声小段落毎に求めた強調確率  $P_s(e)$  と平静確率  $P_s(n)$  などを各音声小段落毎に仕分けして記憶手段に音声強調確率テーブルとして格納する。図15に記憶手段に格納した音声強調確率テーブルの一例を示す。図15に示すF1、F2、F3…は音声小段落毎に求めた音声小段落強調確率  $P_s(e)$  と、音声小段落平静確率  $P_s(n)$  を記録した小段落確率記憶部を示す。これらの小段落確率記憶部F1、F2、F3…には各音声小段落Sに付された音声小段落番号iと、開始時刻(要約対象となる音声データ列の先頭から計時した時刻)終了時刻、音声小段落強調確率、音声小段落平静確率、各音声小段落を構成するフレーム数fn等が格納される。

【0039】要約条件入力ステップS12で入力する条件としては要約すべきコンテンツの全長を  $1/X$  (Xは正の整数)の時間に要約することを示す要約率r(請求の範囲記載の要約率の逆数  $r = 1/X$  を指す)、あるいは要約時間tを入力する。この要約条件の設定に対し、抽出条件変更ステップS13では初期値として重み係数Wを  $W = 1$  に設定し、この重み係数を要約抽出ステップS14に入力する。要約抽出ステップS14は重み係数  $W = 1$  として音声強調確率テーブルから各音声小段落毎に格納されている強調確率  $P_s(e)$  と平静確率  $P_s(n)$  とを比較し、

$$W \cdot P_s(e) > P_s(n)$$

の関係にある音声小段落を抽出すると共に、更にこの抽出した音声小段落を一つでも含む音声段落を抽出し、抽出した音声段落列の総延長時間MT(分)を求める。

【0040】抽出した音声段落列の総延長時間MT(分)と要約条件で決めた所定の要約時間YT(分)とを比較する。ここで  $MT \approx YT$  (YTに対するMTの誤差が例えば±数%程度の範囲)であればそのまま採用した音声段落列を要約音声として再生する。要約条件で設定した要約時間YTに対するコンテンツの要約した総延長時間MTとの誤差値が規定より大きく、その関係が  $MT > YT$  であれば抽出した音声段落列の総延長時間MT(分)が、要約条件で定めた要約時間YT(分)より長いと判定し、図13に示した抽出条件変更ステップS13を再実行させる。抽出条件変更ステップS13では重み係数が  $W = 1$  で抽出した音声段落列の総延長時間MT(分)が要約条件で定めた要約時間YT(分)より「長い」とする判定結果を受けて強調確率  $P_s(e)$  に現在値より小さい重み付け係数W(請求項1記載の所定の係数の場合は現在値よりも大きくする)を乗算  $W \cdot P_s(e)$  して重み付けを施す。重み係数Wとしては例えば  $W = 1 - 0.001 \times K$  (Kはループ回数)で求める。

【0041】つまり、音声強調確率テーブルから読み出した音声段落列の全ての音声小段落で求められている強



調確率 $P_s(e)$ の配列に1回目のループでは $W=1-0.001 \times 1$ で決まる重み係数 $W=0.999$ を乗算し、重み付けを施す。この重み付けされた全ての各音声小段落の強調確率 $W \cdot P_s(e)$ と各音声小段落の平静確率 $P_s(n)$ とを比較し、 $W \cdot P_s(e) > P_s(n)$ の関係にある音声小段落を抽出する。この抽出結果に従って要約抽出ステップS14では抽出された音声小段落を含む音声段落を抽出し、要約音声段落列を再び求める。これと共に、この要約音声段落列の総延長時間MT(分)を算出し、この総延長時間MT(分)と要約条件で定められる要約時間YT(分)とを比較する。比較の結果が $MT \approx YT$ であれば、その音声段落列を要約音声と決定し、再生する。

【0042】1回目の重み付け処理の結果が依然として $MT > YT$ であれば抽出条件変更ステップを、2回目のループとして実行させる。このとき重み係数 $W$ は $W=1-0.001 \times 2$ で求める。全ての強調確率 $P_s(e)$ に $W=0.998$ の重み付けを施す。このように、ループの実行を繰り返す毎にこの例では重み係数 $W$ の値を徐々に小さくするように抽出条件を変更していくことにより $WP_s(e) > P_s(n)$ の条件を満たす音声小段落の数を漸次減らすことができる。これにより要約条件を満たす $MT \approx YT$ の状態を検出することができる。

【0043】尚、上述では要約時間MTの収束条件として $MT \approx YT$ としたが、厳密に $MT = YT$ に収束させることもできる。この場合には要約条件に例えば5秒不足している場合、あと1つの音声段落を加えると10秒超過してしまうが、音声段落から5秒のみ再生することで利用者の要約条件に一致させることができる。また、この5秒は強調と判定された音声小段落の付近の5秒でもよいし、音声段落の先頭から5秒でもよい。また、上述した初期状態で $MT < YT$ と判定された場合は重み係数 $W$ を現在値よりも小さく例えば $W=1-0.001 \times K$ として求め、この重み係数 $W$ を平静確率 $P_s(n)$ の配列に乗算し、平静確率 $P_s(n)$ に重み付けを施せばよい。また、他の方法としては初期状態で $MT > YT$ と判定された場合に重み係数を現在値より大きく $W=1+0.001 \times K$ とし、この重み係数 $W$ を平静確率 $P_s(n)$ の配列に乗算してもよい。

【0044】また、要約再生ステップS15では要約抽出ステップS14で抽出した音声段落列を再生するものとして説明したが、音声付の画像情報の場合、要約音声として抽出した音声段落に対応した画像情報を切り出してつなぎ合わせ、音声と共に再生することによりテレビ放送の要約、あるいは映画の要約等を行うことができる。また、上述では音声強調確率テーブルに格納した各音声小段落毎に求めた強調確率又は平静確率のいずれか一方に直接重み係数 $W$ を乗算して重み付けを施すことを説明したが、強調状態を精度良く検出するためには重み係数 $W$ に各音声小段落を構成するフレームの数 $F$ 乗して

WFとして重み付けを行うことが望ましい。

【0045】つまり、式(8)及び式(10)で算出する条件付の強調確率 $P_s(e)$ は各フレーム毎に求めた強調状態となる確率の積を求めている。また平静状態となる確率 $P_s(n)$ も各フレーム毎に算出した平静状態となる確率の積を求めている。従って、例えば強調確率 $P_s(e)$ に重み付けを施すには各フレーム毎に求めた強調状態となる確率毎に重み付け係数 $W$ を乗算すれば正しい重み付けを施したことになる。この場合には音声小段落を構成するフレーム数を $F$ とすれば重み係数 $W$ は $WF$ となる。この結果、フレームの数 $F$ に応じて重み付けの影響が増減され、フレーム数の多い音声小段落ほど、つまり延長時間が長い音声小段落程大きい重みが付されることになる。

【0046】但し、単に強調状態を判定するための抽出条件を変更すればよいのであれば各フレーム毎に求めた強調状態となる確率の積又は平静状態となる積に重み係数 $W$ を乗算するだけでも抽出条件の変更を行うことができる。従って、必ずしも重み付け係数 $W$ を $WF$ とする必要はない。また、上述では抽出条件の変更手段として音声小段落毎に求めた強調確率 $P_s(e)$ 又は平静確率 $P_s(n)$ に重み付けを施して $P_s(e) > P_s(n)$ を満たす音声小段落の数を変化させる方法を採用したが、他の方法として全ての音声小段落の強調確率 $P_s(e)$ と平静確率 $P_s(n)$ に関してその確率比 $P_s(e)/P_s(n)$ を演算し、この確率比の降順に対応する音声信号区間(音声小段落)を累積して要約区間の和を算出し、要約区間の時間の総和が、略所定の要約時間に合致する場合、そのときの音声信号区間を要約区間と決定して要約音声を編成する方法も考えられる。

【0047】この場合、編成した要約音声の総延長時間が要約条件で設定した要約時間に対して過不足が生じた場合には、強調状態にあると判定するための確率比 $P_s(e)/P_s(n)$ の値を選択する閾値を変更すれば抽出条件を変更することができる。この抽出条件変更方法を採用場合には要約条件を満たす要約音声を編成するまでの処理を簡素化することができる利点が得られる。上述では各音声小段落毎に求める強調確率 $P_s(e)$ と平静確率 $P_s(n)$ を各フレーム毎に算出した強調状態となる確率の積及び平静状態となる確率の積で算出するものとして説明したが、他の方法として各フレーム毎に求めた強調状態となる確率の平均値を求め、この平均値をその音声小段落の強調確率 $P_s(e)$ 及び平静確率 $P_s(n)$ として用いることもできる。

【0048】従って、この強調確率 $P_s(e)$ 及び平静確率 $P_s(n)$ の算出方法を採用場合には重み付けに用いる重み付け係数 $W$ はそのまま強調確率 $P_s(e)$ 又は平静確率 $P_s(n)$ に乗算すればよい。図16を用いて要約率を自由に設定することができる音声処理装置の実施例を示す。この実施例では図12に示した音声強調状

態要約装置の構成に要約条件入力部 31 と、音声強調確率テーブル 32 と、強調小段落抽出部 33 と、抽出条件変更部 34 と、要約区間仮判定部 35 と、この要約区間仮判定部 35 の内部に要約音声の総延長時間を求める総延長時間算出部 35A と、この総延長時間算出部 35A が算出した要約音声の総延長時間が要約条件入力部 31 で入力した要約時間の設定の範囲に入っているか否かを判定する要約区間決定部 35B と、要約条件に合致した要約音声を保存し、再生する要約音声保存・再生部 35C を設けた構成とした点を特徴とするものである。

【0049】入力音声は図 11 で説明したように、フレーム毎に音声特徴量が求められ、この音声特徴量に従って強調確率計算部 16 と平静確率計算部 17 でフレーム毎に強調確率と、平静確率とを算出し、これら強調確率と平静確率を各フレームに付与したフレーム番号と共に記憶部 12 に格納する。更に、このフレーム列番号に音声小段落判定部で判定した音声小段落列に付与した音声小段落列番号が付記され、各フレーム及び音声小段落にアドレスが付与される。この実施例に示した音声処理装置では強調確率算出部 16 と平静確率算出部 17 は記憶部 12 に格納している各フレームの強調確率と平静確率を読み出し、この強調確率及び平静確率から各音声小段落毎に強調確率  $P_s(e)$  と平静確率  $P_s(n)$  とを求め、これら強調確率  $P_s(e)$  と平静確率  $P_s(n)$  を音声強調テーブル 32 に格納する。

【0050】音声強調テーブル 32 には各種のコンテンツの音声波形の音声小段落毎に求めた強調確率と平静確率とが格納され、いつでも利用者の要求に応じて要約が実行できる体制が整えられている。利用者は要約条件入力部 31 に要約条件を入力する。ここで言う要約条件とは要約したいコンテンツの名称と、そのコンテンツの全長時間に対する要約率を指す。要約率としてはコンテンツの全長を  $1/10$  に要約するか、或は時間で 10 分に要約するなどの入力方法が考えられる。ここで例えば  $1/10$  と入力した場合は要約時間算出部 31A はコンテンツの全長時間を  $1/10$  した時間を算出し、その算出した要約時間を要約区間仮判定部 35 の要約区間決定部 35B に送り込む。

【0051】要約条件入力部 31 に要約条件が入力されたことを受けて制御部 19 は要約音声の生成動作を開始する。その開始の作業としては音声強調テーブル 32 から利用者が希望したコンテンツの強調確率と平静確率を読み出す。読み出された強調確率と平静確率を強調小段落抽出部 33 に送り込み、強調状態にあると判定される音声小段落番号を抽出する。強調状態にある音声区間を抽出するための条件を変更する方法としては上述した強調確率  $P_s(e)$  又は平静確率  $P_s(n)$  に確率比の逆数となる重み付け係数  $W$  を乗算し  $W \cdot P_s(e) > P_s(n)$  の関係にある音声小段落を抽出し、音声小段落を含む音声段落により要約音声を得る方法と、確率比  $P_s$

( $e$ ) /  $P_s(n)$  を算出し、この確率比を降順に累算して要約時間を得る方法とを用いることができる。

【0052】抽出条件の初期値としては重み付けにより抽出条件を変更する場合には重み付け係数  $W$  を  $W=1$  としして初期値とすることが考えられる。また、各音声小段落毎に求めた強調確率  $P_s(e)$  と平静確率  $P_s(n)$  の確率比  $P_s(e) / P_s(n)$  の値に応じて強調状態と判定する場合は初期値としてその比の値が例えば  $P_s(e) / P_s(n) \geq 1$  である場合を強調状態と判定することが考えられる。この初期設定状態で強調状態と判定された音声小段落番号と開始時刻、終了時刻を表わすデータを強調小段落抽出部 33 から要約区間仮判定部 35 に送り込む。要約区間仮判定部 35 では強調状態と判定した小段落番号を含む音声段落を記憶部 12 に格納している音声段落列から検索し、抽出する。抽出した音声段落列の総延長時間を総延長時間算出部 35A で算出し、その総延長時間と要約条件で入力された要約時間とを要約区間決定部 35B で比較する。比較の結果が要約条件を満たしていれば、その音声段落列を要約音声保存・再生部 35C で保存し、再生する。この再生動作は強調小段落抽出部 33 で強調状態と判定された音声小段落の番号から音声段落を抽出し、その音声段落の開始時刻と終了時刻の指定により各コンテンツの音声データ或は映像データを読み出して要約音声及び要約映像データとして送出する。

【0053】要約区間決定部 35B で要約条件を満たしていないと判定した場合は、要約区間決定部 35B から抽出条件変更部 34 に抽出条件の変更指令を出力し、抽出条件変更部 34 に抽出条件の変更を行わせる。抽出条件変更部 34 は抽出条件の変更を行い、その抽出条件を強調小段落抽出部 33 に入力する。強調小段落抽出部 33 は抽出条件変更部 34 から入力された抽出条件に従って再び音声強調確率テーブル 32 に格納されている各音声小段落の強調確率と平静確率との比較判定を行う。強調小段落抽出部 33 の抽出結果は再び要約区間仮判定部 35 に送り込まれ、強調状態と判定された音声小段落を含む音声段落の抽出を行わせる。この抽出された音声段落の総延長時間を算出し、その算出結果が要約条件を満たすか否かを要約区間決定部 35B で行う。この動作が要約条件を満たすまで繰り返され、要約条件が満たされた音声段落列が要約音声及び要約映像データとして記憶部 12 から読み出されユーザ端末に配信される。

【0054】尚、上述では要約区間の開始時刻及び終了時刻を要約区間と判定した音声段落列の開始時刻及び終了時刻として取り出すことを説明したが、映像付のコンテンツの場合は要約区間と判定した音声段落列の開始時刻と終了時刻に接近した映像信号のカット点を検出し、このカット点（例えば特開平 8-32924 号公報記載のように検出した画面の切替わりに発生する信号を利用する）の時刻で要約区間の開始時刻及び終了時刻を規定



する方法も考えられる。このように映像信号のカット点を要約区間の開始時刻及び終了時刻に利用した場合は、要約区間の切替わりが画像の切替わりに同期するため、視覚上で視認性が高まり要約の理解度を向上できる利点が得られる。

【0055】上述したように、この発明で用いる音声要約方法は音声波形の強調状態となる確率が高い音声区間を要約区間として抽出するから、この音声区間と同一の開始時間と終了時間で抽出される映像情報もコンテンツの内容で重要な映像部分である場合が多い。この結果、この発明で用いる要約方法によればコンテンツの内容を適格に視聴者に伝えることができる要約情報を得ることができる利点が得られる。図1にこの発明による映像再生方法の処理手順の一例を示す。この発明では放送中の実時間映像信号と音声信号を再生時刻と対応付けて録画し、この録画状態を維持しながら、録画されている部分を要約再生しようとするものである。以下この状況を追いつき再生と称することにする。

【0056】ステップS111では追いつき再生対象開始時刻又は追いつき再生対象開始映像の特定を行う。この特定のためには例えば視聴者が番組視聴中に一時離席する場合に、押ボタン操作により離席時刻の指定を行う。又は室のドアにセンサが設けられ、ドアの開閉に伴って視聴者が室外に退出したことを検知して離席時刻を指定する。又は離席とは関係なく、既に録画されている番組の一部を早送り再生し、その映像から視聴者が任意に追いつき再生の開始位置を特定するなどの方法が考えられる。ステップS112では要約条件（要約時間又は要約率）の入力を行う。この入力とは離席した視聴者が席に戻った時点で行われる。つまり、離席中の時間が例えば30分間であった場合に、視聴者はその30分間に放送された内容をどの程度に圧縮して視聴するかを視聴者の考えに従って要約条件を入力する。その他の入力方法としては視聴者からの入力が無かった場合に予め定められたデフォルト値例えば3分間を使用するか、又は幾つかの候補を用意しておき、これらを表示し、視聴者が指示選択して入力する方法が考えられる。

【0057】また、例えば予約録画により自動的に録画が開始されている状態で視聴者が帰宅した場合、予約録画により録画開始時刻が既知であるため、要約再生開始の指示により要約終了時刻が決定される。ここで例えば要約条件が予めデフォルト値等により決定されていれば、その要約条件に従って録画開始時刻から要約終了時刻までが要約される。ステップS113で追いつき再生開始を指定する。この追いつき再生開始の指定により、要約対象区間の終了点（要約終了時刻）が指定される。入力方法としては視聴者が追いつき再生開始を指示する押ボタン操作によりその操作時刻を指定するか、又は室のドアに設けた開閉センサにより視聴者の入室を検出し、入室時刻を持って追いつき再生開始の指定を行って

もよい。

【0058】ステップS114で現在放送中の映像再生を停止する。ステップS115で要約処理、要約映像・音声の再生を行う。要約処理はステップS111で定めた追いつき再生対象開始時刻又は追いつき再生対象開始映像からステップS113で入力した要約終了時刻までの音声信号について指定された要約条件に従って、要約区間を特定し、その要約区間の音声信号と、要約区間と同期する映像信号を再生する。ステップS116で要約区間の再生が終了。ステップS117で放送中の映像再生を再開する。

【0059】図2に上述した追いつき再生を実行するこの発明の映像再生装置100の一例を示す。この発明による映像再生装置100は、記録部101と、音声分離部102と、音声要約部103と、要約区間読出部104と、モード切替部105とによって構成することができる。記録部101は、例えばハードディスク或は半導体メモリ、DVD-ROM等のように高速で書込および読出を実行することができる記録再生手段を用いる。高速で書込及び読出を実行することにより、現在放送中の番組を録画しながら、既に記録されている部分を再生することができる。入力信号S1はテレビチューナー等から入力され、アナログ信号でもデジタル信号でもどちらでもよい。但し、記録部101の記録はデジタル信号で行われる。

【0060】音声分離部102は指定された要約対象区間の映像信号から音声を分離し、この音声信号を音声要約部103に入力する。音声要約部103はこの音声信号を用いて強調部分を抽出し、要約区間を特定する。尚、音声要約部103は録画中は常に音声信号を分析し、録画している番組毎に図15に示した音声強調テーブルを作成し記憶部に格納する。従って、番組の放映中に途中から録画部分の要約再生を行う場合は音声強調テーブルを用いて要約が行われる。また後日、録画された番組の要約を視聴する場合も音声強調確率テーブルを用いて要約が行われる。

【0061】要約区間読出部104は音声要約部103で特定された要約区間に従って記録部101から音声付映像信号を読み出し、モード切替部105に出力する。モード切替部105は要約区間読出部104が読み出した音声付映像信号を要約映像信号として出力し、視聴者の視聴に提供される。モード切替部105は要約映像を出力するモードaの他に、記録部101から読み出した映像信号を出力する再生モードbと、入力信号S1を直接視聴に提供するモードcとに切替えられ、各種の形態で利用が可能とされる。ところで、上述した追いつき再生方法には追いつき再生の実行期間中に放送された映像は要約対象区間に含まれないため、視聴者にはその映像を視聴することができない不都合が生じる。

【0062】このため、この発明ではこの欠点を解消す

るために、以下の実施例を提案する。つまり、要約区間の再生が終了後に、以前の再生開始時刻を新たな要約対象開始時刻とし、要約区間再生終了時刻を新たな要約終了時刻として要約処理及び要約映像・音声の再生処理を繰り返す。以前の再生開始時刻と要約区間再生終了時刻の間の時間が所定時間（例えば5～10秒）以下となる場合に繰り返しを終了する。この場合は指定された要約率もしくは要約時間よりも長く要約区間が再生されるとい問題が生じる。例えば要約対象区間の時間を $T_A$ として要約率 $r$ （ $0 < r < 1$ 、 $r$  = 要約時間（要約区間の時間の総延長）／要約対象区間の時間）で要約すると、1回目の要約における要約時間 $T_1$ は $T_A \cdot r$ となる。2回目の要約における要約時間は1回目のものについて更に要約率で要約するので $T_A \cdot r^2$ となる。この処理が順次繰り返されるので繰り返し終了までに要する時間は $T_A \cdot r / (1 - r)$ となる。

【0063】ここで指定された要約率 $r$ を $r / (1 + r)$ と調整し、この調整された要約率をもって要約を行う。その場合、繰り返し終了までに要する時間は $T_A \cdot r$ となり指定された要約率に適した要約時間となる。同様に要約時間 $T_1$ が指定されたときでも要約対象区間の時間 $T_A$ が与えられていれば、指定された要約率 $r$ は $T_1 / T_A$ であるので、この要約率を $T_1 / (T_A + T_1)$ としても1回目の要約時間を $T_A \cdot T_1 / (T_A + T_1)$ と調整してもよい。図3は上記不都合を解消するための他の実施例を示す。この実施例では入力信号 $S_1$ をそのまま出力させ、表示器の親画面200（図4参照）に現在放送中の映像を表示させると共に、モード切替部105に子画面化処理部106を設け、子画面化処理した要約映像・音声を入力信号 $S_1$ に重畳させて出力し、この要約映像を子画面201（図4参照）に表示させる混成モード $d$ を設けた実施例を示す。

【0064】この実施例によれば、視聴者は親画面200に表示される放映中の番組の内容を視聴しながら、過去に放映された番組の内容を要約して子画面201で視聴することができる。この結果、要約情報を視聴している間に放映されている番組の内容は親画面200から受け取ることができるから、要約情報が全て再生された時点では番組の内容は前半部分から現在放映中の時点までの内容をほぼ切れ目なく視聴者に理解させることができる。以上説明したこの発明による映像再生方法はコンピュータに映像再生プログラムを実行させて実現される。この場合、映像再生プログラムを通信回線を介してダウンロードしたり、CD-ROMや磁気ディスク等の記録媒体に格納させてコンピュータにインストールし、コンピュータ内のCPU等の処理装置で本発明の方法を実行させる。

【0065】

【発明の効果】以上説明したように、この発明によれば録画されている番組を任意の圧縮率に圧縮して要約し、

この要約された要約情報を再生することができる。従って録画されている多数の番組の内容を短時間に見分けることができ、例えば真に視聴したい番組を探す作業を短時間に済ませることができる利点が得られる。更に、最初から視聴することができなかった番組が放映されている状況でも、その前半部分を要約して視聴することができるため、途中から視聴を始めた番組でも、番組の全体を把握して視聴を続けることができる。

【図面の簡単な説明】

【図1】この発明による映像再生方法を実行する場合の手順の一例を説明するためのフローチャート。

【図2】この発明による映像再生装置の一実施例を説明するためのブロック図。

【図3】この発明による映像再生装置の変形実施例を説明するためのブロック図。

【図4】この発明による映像再生装置の特徴を説明するための表示画面の一例を説明するための図。

【図5】先に提案した音声要約方法を説明するためのフローチャート。

【図6】先に提案した音声段落の抽出方法を説明するためのフローチャート。

【図7】音声段落と音声小段落の関係を説明するための図。

【図8】図5に示したステップS2における入力音声小段落の発話状態を判定する方法の例を示すフローチャート。

【図9】先に提案した音声要約方法に用いられるコードブックを作成する手順の例を示すフローチャート。

【図10】この発明において用いられるコードブックの記憶例を示す図。

【図11】発話状態尤度計算を説明するための波形図。

【図12】先に提案した音声強調状態判定装置及び音声要約装置の一実施例を説明するためのブロック図。

【図13】要約率を自由に変更することができる要約方法を説明するためのフローチャート。

【図14】音声の要約に用いる音声小段落の抽出動作と各音声小段落の強調確率算出動作、音声小段落平静確率抽出動作を説明するためのフローチャート。

【図15】音声要約装置に用いる音声強調確率テーブルの構成を説明するための図。

【図16】要約率を自由に変更することができる音声要約装置の一例を説明するためのブロック図。

【符号の説明】

100 映像再生装置

101 記録部

102 音声

分離部

103 音声要約部

104 要約

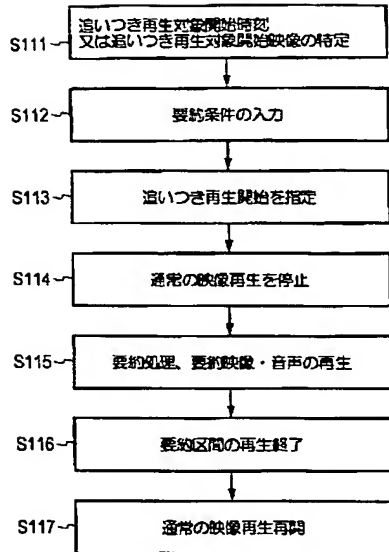
区間読出部

105 モード切替部

106 子画

面化処理部

【図1】



【図4】

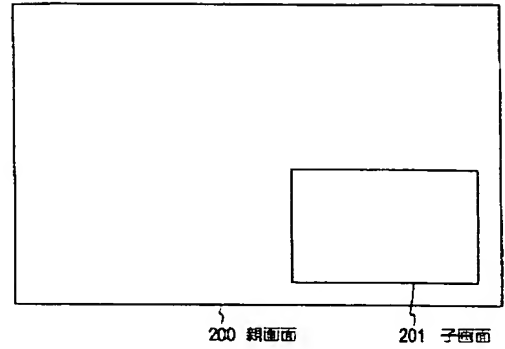


図 4

図 1

【図2】

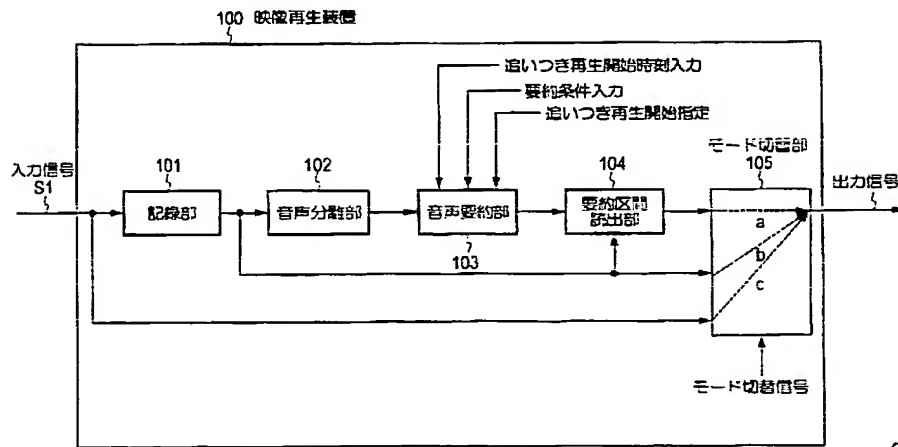


図 2

【図3】

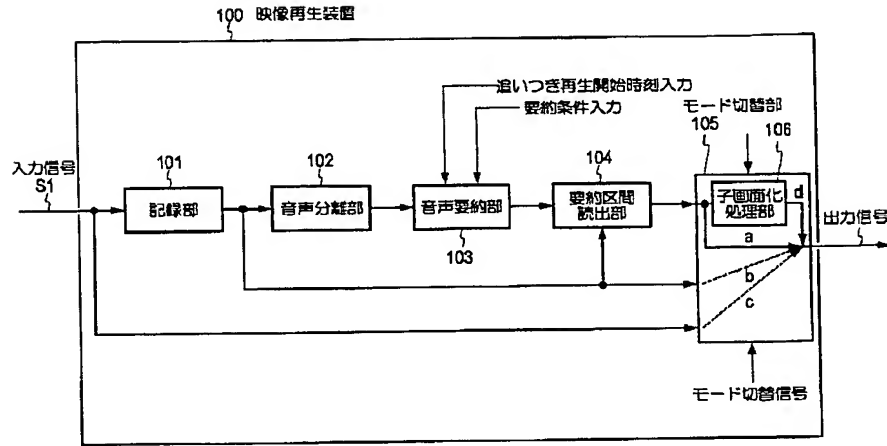


図 3

【図5】

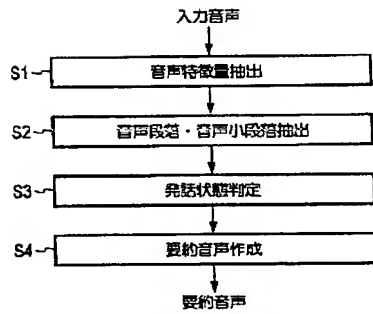


図 5

【図6】

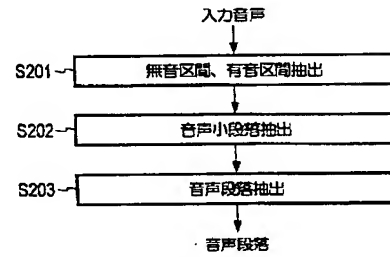


図 6

【図7】

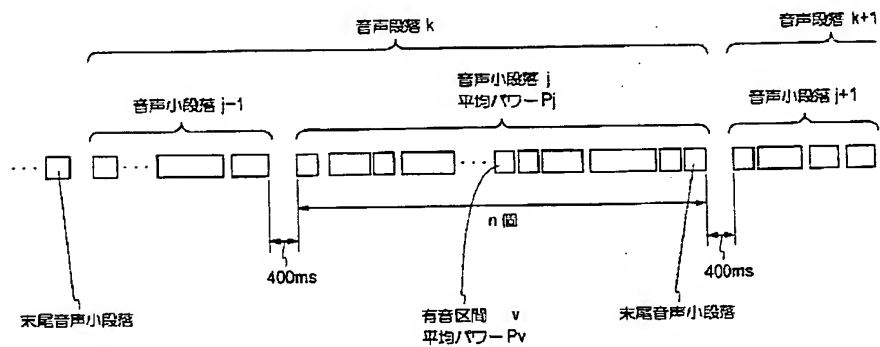


図 7

【図8】

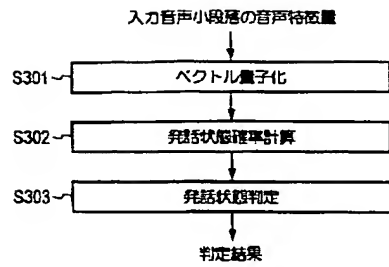


図 8

【図9】

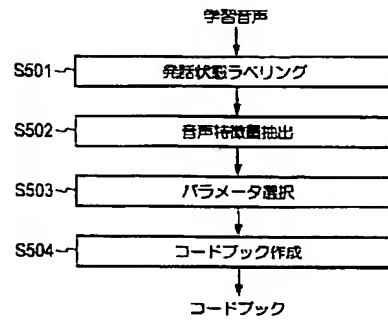


図 9

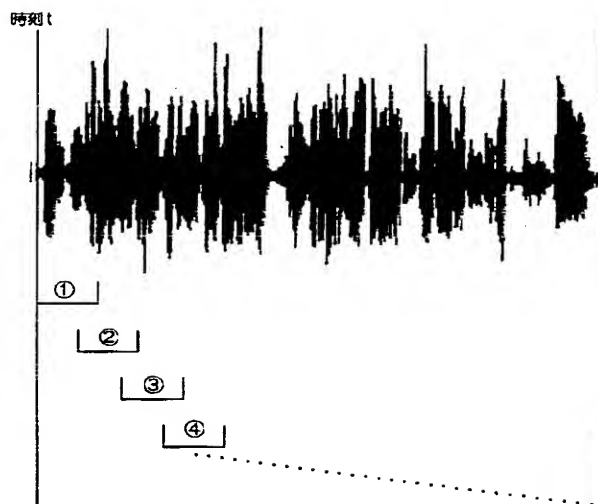
【図10】

コードブック

		単独出現確率		条件付確率											
コード No.	音素 特徴量			C1		C2		C3		C2, C3		C3, C4			
		強調	平静	強調	平静	強調	平静	強調	平静	強調	平静	強調	平静		
C1															
C2															
C3															
⋮	⋮														
⋮	⋮														

図 10

【図11】



【図13】

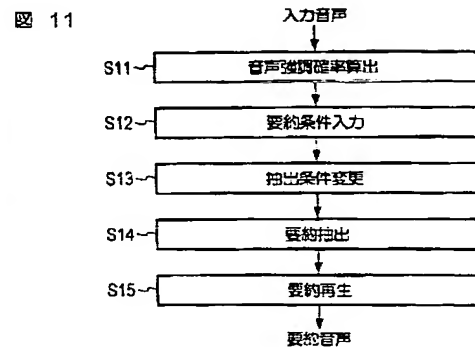
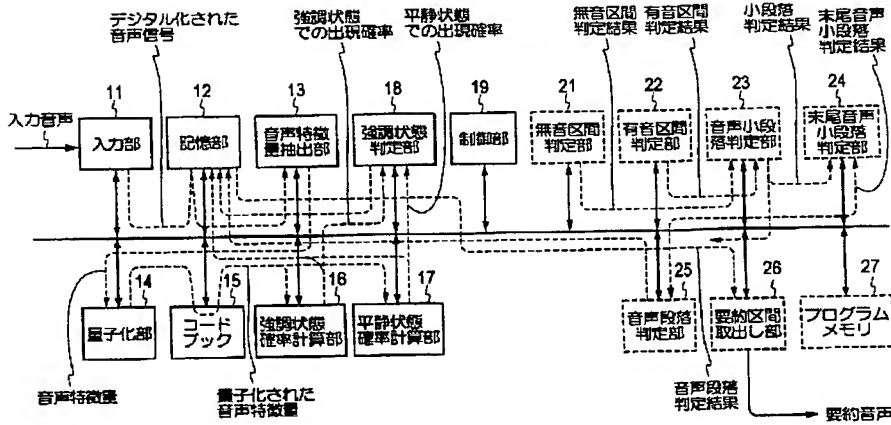


図 13

【図12】

図 12



【図14】

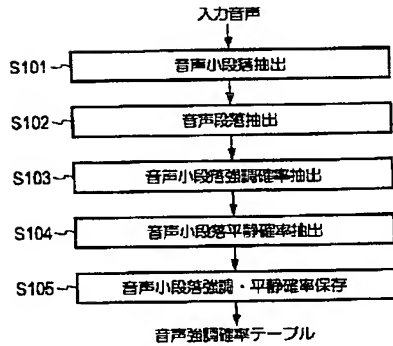


図 14

【図15】

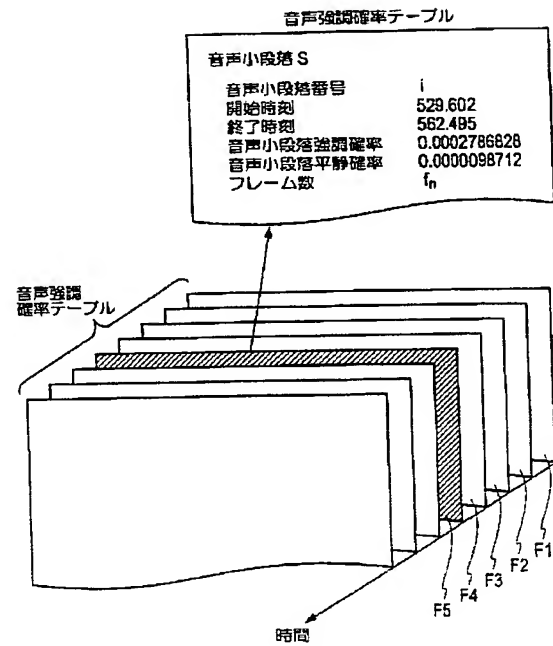


図 15

[illegible]

(51) Int. Cl. <sup>7</sup>	識別記号	F I	テーマコード (参考)
G 1 0 L 15/06		H 0 4 N 5/91	C
15/10		G 1 0 L 3/00	5 3 1 N
H 0 4 N 5/91			5 2 1 U
			5 1 3 A

(72)発明者 水野 理  
東京都千代田区大手町二丁目 3 番 1 号 日  
本電信電話株式会社内

F ターム(参考) 5C053 HA27 JA23 LA11  
5D015 DD03 FF00 FF05 HH04 KK01

THIS PAGE BLANK (USPTO)



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)